

Network Monitoring

Joe Metzger

**Energy Sciences Network
Lawrence Berkeley National Laboratory**

**May 14 2008
US LHC T1/T2/T3 Networking Workshop**

Networking for the Future of Science



Current Network Environment

- Most R&E network backbones are composed of 10Gbps links
- The LHC community has the tools, techniques, infrastructure & capability to transfer data at 10Gbps.
- But...
 - Network topology is **constantly** changing!
 - LHC data transfer flows are not typical internet flows
 - Many network operators don't have a lot of experience with large flows
 - Most physics flows cross multiple domains
 - Many cross-domain links haven't been tested at capacity
 - Line rate flows don't aggregate nicely
 - Debugging problems can be difficult

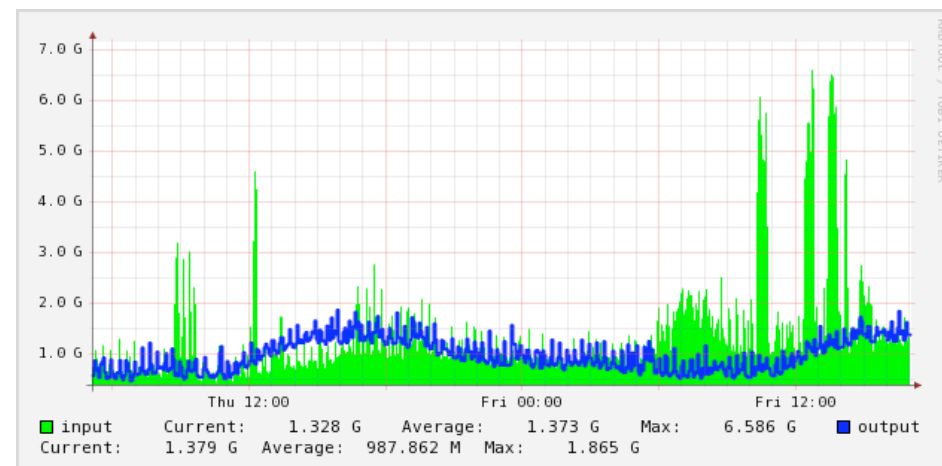
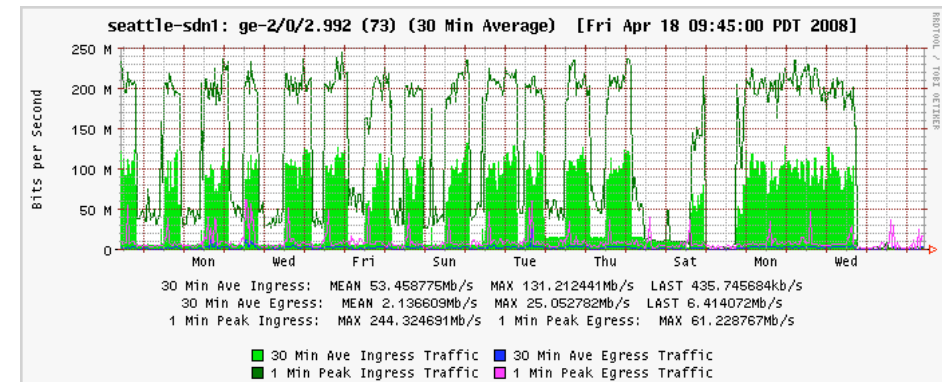
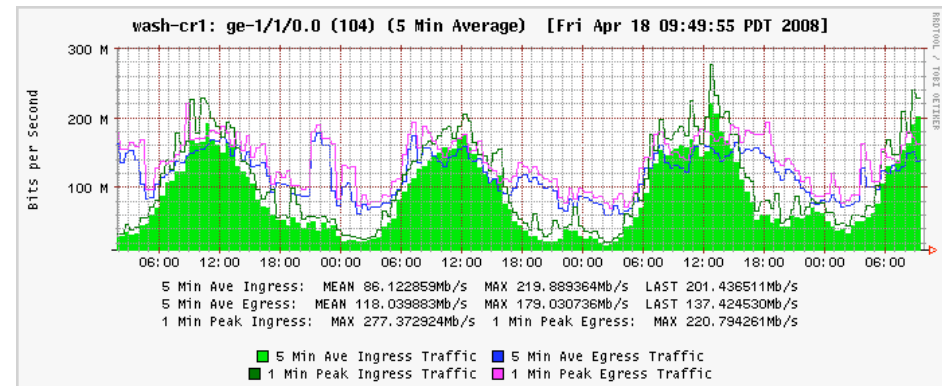
Measurement Requirements

- You must have the ability to easily determine the status of the set of paths you rely on for your critical missions.
 - Up and working correctly?
 - **How do you prove it?**
 - Down
 - Is there a known problem that is being worked on?
 - Are you seeing a symptom of the problem or something else?
 - Is part of the network down or the applications down?
 - How do you prove the problem is, or is not in your cluster/campus/regional?
 - Who do you call and **what hard data can you provide to help them quickly identify the problem and fix it?**
 - Up but not performing as expected.
 - Is there a known problem?
 - Who do you call and **what hard data can you provide to help them quickly identify the problem and fix it?**
- You need to understand the amount network resources you are consuming.
 - Are you getting more, less, or approximately your fair share?
 - Do you know if/how your use of the network is affecting others?

New Network Traffic Profiles

- Old Typical Traffic Pattern
- Steady State Instrument Output Pattern
- Tuned Bulk Data Transfer

How many 5-7 Gbps flows can you aggregate on a 10 Gbps backbone?



Community Progress

- There has been a lot of work in the Network Measurement space
 - Developing frameworks for exchanging measurement data
 - Developing & improving measurement tools
 - Defining diagnostic methodologies
 - Analysis techniques
- There is a small community that understands how to use these tools and techniques for network performance analysis, verification and debugging
- The LHC community should take advantage of these capabilities
 - Metcalfe's law - The value of a network is proportional to the square of the number of users.

LHC US Tier 1/2/3 Measurement Document(s)

- Introduction
 - A general discussion of the value of network measurement to the community.
- Best Practices
 - What Measurements to support
 - Delay, Bandwidth, Interface Utilization, Errors & Discards, etc
 - Protocols
 - For measurement collection: ICMP, OWAMP, iperf etc.
 - For measurement Publication & Sharing - perfSONAR
 - Schedules & parameters
 - For regularly scheduled tests
 - Data sharing guidelines
- Implementation Guide
 - What tools to use
 - How to configure them
- Usage Guide
 - How to use the Measurement Infrastructure to:
 - Verify important network infrastructure is running properly
 - Debug problems efficiently

Technical Requirements Latency

1. Continuously measure end-to-end **delay**

A. What

- Run continuous tests and store results in an MA
- Publish results via a standardized web service interface
- Provide a tool to visualize the data
- Provide tools to automatically analyze data and generate NOC alarms

B. Why

- Measure & document actual availability
- Provide time references for when problems occurred and when they were fixed
- Detect & assist in diagnosing common causes of performance degradation
 - a. Packet Loss
 - Congestion related
 - Non-Congestion related
 - b. Queuing & Jitter caused by congestion
 - c. Routing Issues: changes, asymmetry, flapping, etc

Technical Requirements Bandwidth

2. Make regular scheduled **bandwidth** measurements across paths of interest

A. What

- Run regularly scheduled tests and store results in an MA
- Publish results via a standardized web service interface
- Provide a tool to visualize the data
- Provide tools to automatically analyze data and generate NOC alarms

B. Why

- Detect performance problems
- Identify when problems appeared
- Document performance delivered

Technical Requirements Circuit Status

3. Monitor up/down status of cross domain circuits

A. What

- Determine the status of a circuit
- Publish status via a web services interface
- Provide tools to visualize state
- Generate NOC alarms when circuits change states

B. Why

- Determine when circuits are available
- Simplify debugging of end to end circuit problems

Technical Requirements Interface Statistics

4. Monitor Link/Circuit Capacity, Utilization & Errors

A. What

- Publish statistics via a web services interface
- Provide tools to visualize the data
- Generate NOC alarms when thresholds are crosses

B. Why

- Allow determining usage patterns
- Simplify throughput problem diagnosis
- Capacity Planning

Technical Requirements Topology

5. Measure & Publish Topology of primary and backup paths

A. What

- Publish logical topology via a web services interface
- Provide tools to visualize the data over time

B. Why

- Set user expectations
- Facilitate network problem diagnosis
- Allow correlating logical topology to measurements of the physical topology
- Understand ...

Implementation Details Latency

1. Continuously measure end-to-end **delay**

A. Tools

- OWAMP/perfSONARBOUY
- Pinger

B. Configuration

- Each Domain deploys a Measurement Point at the edge of their domain
 - OWAMP + Pinger
- Deploy Scheduler & MA
- Run OWAMP tests to remote sites that have an OWAMP server
- Run Pinger tests to remote sites that are not able or willing to maintain stable Owamp MPs.

Implementation Details Bandwidth

2. Make regular scheduled **bandwidth** measurements across paths of interest

A. Tools

- BWCTL & BWCTL MP
- perfSONAR-BUOY

B. Configuration

- Deploy 1 GE connected MP at the edge of their domain
- Deploy 1 Scheduler & MA per cluster of MP's
 - a. One for the LHCOPN
 - b. One per Tier 1 that wants to measure their Tier 2 service

Implementation Details Circuit Status

3. Monitor up/down status of cross domain circuits

A. Tools

- Topology + Transform service or E2Emp or SQLMA
- E2Emon

B. Configuration

- Each Domain publishes the status of their portions of cross domain circuits.
- E2Ecu monitors all LHC circuits?
- Any NOC can run E2Emon to monitor the subset of circuits that they have responsibility for

Implementation Details Interface Statistics

4. Monitor Link/Circuit Capacity, Utilization & Errors

A. Tools

- PS-SNMP MA or RRDma

B. Configuration

- Each domain sets up a Measurement Archive publishing statistics about their network interfaces supporting LHC
 - a. Capacity
 - b. Utilization
 - c. Input Errors
 - d. Output Drops

Implementation Details Topology

5. Measure & Publish Topology of primary and backup paths

A. Tools Still Under Development

- Internet2 Topology Service Internet2 Topology Service
- CNIS

This is not a significant concern for the LHCOPN as long as it continues to be a well defined static topology fully described with the E2Emon tools.

This is an issue when considering Tier 2 traffic which will stress the global R&E Networking Infrastructure!

Next Steps

- Identify community representatives to participate in writing & reviewing document
 - Backbone Network Measurement specialists
 - Physics Users
 - Tier 1 center network operators
 - Campus Networking Person
- Present idea & current status at Internet2 Spring Member Meeting HENP-SIG
- Present draft recommendations to the US LHC community at the T1/T2/T3 meeting at BNL in May
 - <https://wiki.internet2.edu/confluence/display/PSPS/Tier-2-BCP>
- Evaluate the 'US Recommendations' applicability to the global environment at LHCOPN meeting in June
- Present recommendations & pilot implementations at Joint Techs in July
- US LHC community using infrastructure before end of summer